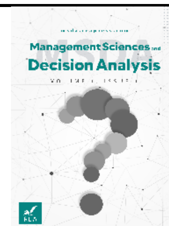
E-ISSN: | P-ISSN:



Management Sciences and Decision Analysis

www.msda.reapress.com

Manag. Sci. Decis. Anal. Vol. 3, No. 2 (2025) 108–123.





Paper Type: Original Article

Portfolio optimization using DEA integration with multiple data sources with machine learning approach

Morteza Sarmadi[1], **Seyyed Ismail Najafi**[2,*], **Seyyed Ahmad Edalatpanah**[3]

[1] *Department of Financial Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran; sarmadi787@gmail.com.*

[2] *Department of Industrial Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran; najafi1515@yahoo.com.*

[3] *Department of Applied Mathematics, Aindegan Institute of Higher Education, Tonekabon, Iran; saedalatpanah@gmail.com.*

Citation:



Sarmadi, M., Najafi, S. I., & Edalatpanah, S. A. (2025). Portfolio optimization using DEA integration with multiple data sources with machine learning approach. *Management sciences and decision analysis*, *3*(2), 108-123.

Received: 01/06/2024 Reviewed: 28/06/2024 Revised: 18/08/2024 Accepted: 22/11/2024

Abstract

Objective: Portfolio optimization is a hot research topic that has attracted the attention of many researchers in recent decades. The existing literature shows that the performance of traditional mean-variance portfolio strategies is not suitable.

Method: To address this issue, this paper first explains that a complete investment process includes two parts: Asset selection and investment weight formulation. Then, we design a stock selection scheme that integrates Data Envelopment Analysis (DEA) with multiple data sources to evaluate the investment value of stocks in terms of historical returns and asset correlation. In addition, we use a Support Vector Machine (SVM) combined with multi-source data to predict stock price movements. In the next step, we combine the obtained stock price movements and the proposed stock selection scheme to build a portfolio optimization model. We also tested the proposed stock selection scheme and investment strategies, in which the Tehran Stock Exchange index components were selected as experimental samples during the 15 months between June 2021 and November 2022.

Results: The empirical results show that the proposed stock selection scheme can effectively improve the performance of all investment strategies. Moreover, the proposed investment strategy outperforms the traditional minimum global variance investment strategy.

Originality: This research presents an innovative approach to portfolio optimization by integrating Data Envelopment Analysis (DEA) and machine learning algorithms based on diverse data (Technical, fundamental, economic, and sentiment). Unlike traditional methods, this hybrid model allows for a more comprehensive assessment of stock performance on the Tehran Stock Exchange and opens a new horizon for intelligent investment decision-making in emerging markets.

Keywords: Data envelopment analysis, Portfolio optimization, Machine learning, Multiple data sources.



Corresponding Author: najafi1515@yahoo.com



https://doi.org/10.22105/msda.v3i2.60





علوم مدیریت و تحلیل تصمیم

دوره ۳، شماره (۲)، (۲۰۲۵)، ۱۲۳-۱۰۸

www.msda.reapress.com





نوع مقاله: پژوهشی

بهینه‌سازی پرتفوی با استفاده از ادغام تحلیل پوششی داده‌ها با منابع داده چندگانه با رویکرد یادگیری ماشین در بورس اوراق بهادار تهران

مرتضی سرمدی[1]، سید اسماعیل نجفی[2]، سید احمد عدالت‌پناه[3]

[1] گروه مهندسی مالی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.

[2] گروه مهندسی صنایع، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.

[3] گروه ریاضی کاربردی، موسسه آموزش عالی آیندگان، تنکابن، ایران.

چکیده

هدف: بهینه‌سازی پرتفوی یک موضوع پژوهشی داغ است که در دهه‌های اخیر توجه بسیاری از پژوهشگران را به خود جلب کرده است. ادبیات موجود نشان می‌دهد که عملکرد استراتژی‌های سنتی پرتفوی میانگین-واریانس مناسب نیست.

روش‌شناسی پژوهش: برای پرداختن به این موضوع، این مقاله ابتدا توضیح می‌دهد که یک فرآیند سرمایه‌گذاری کامل شامل دو بخش انتخاب دارایی و فرمول‌بندی وزن سرمایه‌گذاری است. سپس، یک طرح انتخاب سهام را طراحی می‌کنیم که تحلیل پوششی داده‌ها[1] را با منابع داده چندگانه یکپارچه می‌کند تا ارزش سرمایه‌گذاری سهام را از نظر بازده تاریخی و همبستگی دارایی ارزیابی کند. علاوه بر این، از ماشین بردار پشتیبان[2] ترکیب شده با داده‌های چندمنبعی برای پیش‌بینی حرکات قیمت سهام استفاده می‌کنیم. در گام بعد، از ترکیب حرکات قیمت سهام به‌دست آمده و طرح انتخاب سهام پیشنهادی برای ساخت مدل بهینه‌سازی پرتفوی استفاده می‌کنیم. همچنین یک آزمون بر روی طرح انتخاب سهام پیشنهادی و استراتژی‌های سرمایه‌گذاری انجام می‌دهیم که در آن اجزای شاخص بورس اوراق بهادار تهران در دوره زمانی ۱۵ ماهه بین شهریور ۱۴۰۰ تا آبان ۱۴۰۱ به عنوان نمونه‌های آزمایشی انتخاب می‌شوند.

یافته‌ها: نتایج تجربی نشان می‌دهد که طرح انتخاب سهام پیشنهادی می‌تواند به‌طور مؤثر عملکرد همه استراتژی‌های سرمایه‌گذاری را بهبود بخشد. علاوه بر این، استراتژی سرمایه‌گذاری پیشنهادی در مقایسه با استراتژی سرمایه‌گذاری حداقل واریانس سراسری سنتی عملکرد بهتری دارد.

اصالت/ارزش افزوده علمی: این پژوهش با ادغام تحلیل پوششی داده‌هاو الگوریتم‌های یادگیری ماشین بر پایه داده‌های متنوع (فنی، بنیادی، اقتصادی و احساسی) رویکردی نوآورانه به بهینه‌سازی پرتفوی ارایه می‌دهد. برخلاف روش‌های سنتی، این مدل ترکیبی امکان ارزیابی جامع‌تری از عملکرد سهام در بورس اوراق بهادار تهران فراهم می‌سازد و افق جدیدی برای تصمیم‌گیری هوشمند سرمایه‌گذاری در بازارهای نوظهور می‌گشاید.

کلیدواژه‌ها: تحلیل پوششی داده‌ها، بهینه‌سازی پرتفوی، یادگیری ماشین، منابع داده چندگانه.

۱- مقدمه

بهینه‌سازی پرتفوی یکی از کانون‌های پژوهشی مهم در زمینه مالی از زمان نظریه کلاسیک پرتفوی میانگین-واریانس ارایه شده توسط مارکوویتز بوده است [1]. مدل مارکوویتز به‌عنوان آغاز نظریه مدرن پرتفوی، ابتدا یک راه‌حل کارآمد برای تبادل بین بیشینه‌سازی بازده مورد انتظار و کمینه‌سازی

[1] Data Envelopment Analysis (DEA)

[2] Support Vector Machine (SVM)

ریسک ارایه می‌دهد. این مدل‌های کلاسیک پرتفوی معمولا از میانگین بازده تاریخی سهام به‌عنوان بازده مورد انتظار استفاده می‌کنند که برای سرمایه‌گذاری بلندمدت در بازار سهام مناسب است. اما این مدل‌ها ممکن است برای سرمایه‌گذاری کوتاه‌مدت مناسب نباشند. این پژوهش بر چگونگی تخصیص ثروت سرمایه‌گذاران بین دارایی‌های مختلف و توسعه یک استراتژی سرمایه‌گذاری موثر برای سرمایه‌گذاران، متمرکز است. به بیان دقیق، یک مساله سرمایه‌گذاری کامل شامل دو فرایند انتخاب دارایی و فرمول‌بندی وزن سرمایه‌گذاری است. اولین فرایند شامل تجزیه‌وتحلیل عملکرد جامع دارایی‌های جایگزین (مانند چشم‌انداز صنعت، وضعیت مالی شرکت، عملکرد تاریخی، شرایط بازار و غیره) سپس شناسایی دارایی‌های سرمایه‌گذاری به‌عنوان زیربنای پرتفوی‌ها است.

فرایند دوم را می‌توان به‌عنوان یک مساله بهینه‌سازی (به‌عنوان مثال، به حداقل رساندن ریسک پرتفوی برای یک بازده مورد انتظار معین، یا به حداکثر رساندن بازده مورد انتظار برای یک ریسک پرتفوی معین) که هدف آن تامین سرمایه‌گذاران با وزن سرمایه‌گذاری موثر است، در نظر گرفت. با این حال، مطالعات موجود در مورد بهینه‌سازی پرتفوی عمدتا نقش انتخاب دارایی را نادیده می‌گیرند و فقط مساله انتخاب پرتفوی را به‌عنوان یک مساله بهینه‌سازی خالص در نظر می‌گیرند. در فرایند سرمایه‌گذاری واقعی، انتخاب دارایی‌های اساسی به همان اندازه حیاتی است و بازده پرتفوی به‌دست آمده از انتخاب دارایی‌های مختلف نیز تفاوت‌های زیادی دارد. از طرفی داده‌های چندمنبعی نیز نقش مهمی در تحلیل ارزش سرمایه‌گذاری سهام دارند. با این حال، اکثر مطالعات انجام‌شده از مدل‌های اقتصادسنجی و رویکردهای داده‌کاوی برای تحلیل ارزش سرمایه‌گذاری سهام استفاده می‌کنند؛ بنابراین، نتایج این مطالعات حساسیت بیش‌تری نسبت به تعیین پارامترها و انتخاب مدل‌ها دارد.

تحلیل پوششی داده‌ها، به‌عنوان یک روش ارزیابی ناپارامتریک، می‌تواند به‌طور موثر از چنین مشکلاتی جلوگیری کند، زیرا شکل تابع تولید آن نیازی به مشخص شدن ندارد. رویکرد *DEA* به‌عنوان یک روش ارزیابی مبتنی بر داده بر اساس شاخص‌های ورودی و خروجی چندگانه، به‌طور گسترده در ارزیابی عملکرد واحدهای تصمیم‌گیرنده استفاده شده است. بدیهی است که از این روش می‌توان برای ارزیابی ارزش سرمایه‌گذاری سهام استفاده کرد. همچنین نتایج پیش‌بینی قیمت سهام نیز می‌تواند به‌عنوان یک شاخص مهم برای ساخت یک پرتفوی کارآمد درنظر گرفته شود. با توجه به تعداد بالای سهام‌داران و حجم بالای نقدینگی در بازار سرمایه، آگاهی نسبت به وضعیت آینده سهام به‌منظور دانستن بهترین زمان آن‌جام معامله، چگونگی شناسایی سهام قابل سرمایه‌گذاری در بازار سرمایه با تعداد زیادی دارایی و چگونگی ایجاد یک پرتفوی موثر برای سرمایه‌گذاران و اطمینان خاطر بیش‌تر از موضوعات کلیدی هستند که باید به آن‌ها پرداخته شود. با توجه به موارد فوق هدف اصلی این پژوهش بهینه‌سازی پرتفوی با استفاده از ادغام *DEA* با منابع داده چندگانه با رویکرد یادگیری ماشین خواهد بود. به‌منظور رسیدن به هدف مطرح شده، ابتدا چندین منبع داده (مانند داده‌های تاریخی معاملات، شاخص‌های فنی) را برای تجزیه‌وتحلیل ارزش سرمایه‌گذاری سهام جمع‌آوری می‌کنیم. از *DEA* برای طراحی یک طرح انتخاب سهام استفاده می‌کنیم که ارزش سرمایه‌گذاری سهام را از نظر بازده تاریخی و همبستگی دارایی ارزیابی می‌کند. همچنین، از *SVM* ترکیب شده با داده‌های چندمنبعی برای پیش‌بینی حرکت قیمت سهام استفاده می‌کنیم. علاوه بر این، طرح انتخاب سهام پیشنهادی و حرکات قیمت سهام به‌دست‌آمده را برای ساخت مدل بهینه‌سازی پرتفوی ترکیب می‌کنیم و سپس یک استراتژی سرمایه‌گذاری موثر را برای سرمایه‌گذاران ارایه می‌کنیم.

۲- مبانی نظری و پیشینه پژوهش

مارکوویتز پیشنهاد کرد که "تخم مرغ‌های خود را در یک سبد قرار ندهید [2]"، یعنی سرمایه‌گذاران باید سرمایه‌گذاری‌های خود را تا حد امکان متنوع کنند تا ریسک سرمایه‌گذاری را کاهش دهند. با این حال، در فرایند سرمایه‌گذاری واقعی، اگر پرتفوی بیش از حد متنوع باشد، سرمایه‌گذاران باید هزینه‌های تراکنش بیش‌تری را بپردازند و استراتژی سرمایه‌گذاری پیشنهادی نیز لزوما در عمل بهینه نیست. برای حل این مشکل، کینز و همکاران استدلال کردند که سرمایه‌گذاران باید ثروت خود را در تعداد کمی از سهام با ارزش سرمایه‌گذاری، متمرکز کنند و در نتیجه تنوع پرتفوی‌ها را کاهش دهند [3]. این استراتژی سرمایه‌گذاری به‌عنوان استراتژی کم‌تنوع نیز نامیده می‌شود. از آن زمان، مساله عدم تنوع پرتفوی‌ها به‌طور گسترده مانند [4-8] مورد مطالعه قرار گرفته است. با این حال، مطالعات فوق به چگونگی انتخاب دارایی‌های با ارزش سرمایه‌گذاری نمی‌پردازند؛ بنابراین، چگونگی شناسایی سهام قابل سرمایه‌گذاری در بازار اوراق بهادار با تعداد زیادی دارایی و چگونگی ایجاد یک پرتفوی موثر برای سرمایه‌گذاران نیز موضوعات کلیدی هستند که در این مقاله باید به آن‌ها پرداخته شود.

دلایل انتخاب تحلیل پوششی داده‌ها به‌عنوان الگویی مناسب در اندازه‌گیری کارایی فنی واحدها را می‌توان در توانایی منحصر به فرد این الگو در ارزیابی واقع‌بینانه، ارزیابی هم‌زمان مجموعه عوامل، عدم نیاز به وزن‌های از قبل تعیین شده، جبرانی بودن، ارزیابی با گرایش مرزی و تصویر کردن

بهترین وضعیت عملکردی به‌جای وضعیت مطلوب عنوان کرد [9]. علاوه بر مزایای یاد شده، این روش قادر است کارایی را برای تعداد زیادی از سهام‌های بورس اوراق‌بهادار تهران مستقل از واحد اندازه‌گیری متغیرهای ورودی و خروجی اندازه‌گیری کند؛ بنابراین، در پژوهش حاضر روش ناپارامتریک تحلیل پوششی داده‌ها استفاده شده است.

ادیرایسینگه و ژانگ یک مدل *DEA* تعمیم یافته را برای تجزیه‌وتحلیل اصول بنیادی شرکت‌ها، انتخاب سهام مناسب برای سرمایه‌گذاری و ارایه یک استراتژی سرمایه‌گذاری موثر به سرمایه‌گذاران پیشنهاد کردند [10]. چن از دو مدل *DEA* برای ارزیابی کارایی شرکت‌ها و ایجاد پرتفوی با انتخاب سهام با کارایی بالاتر استفاده کرد. نتایج تجربی نشان داد که پرتفوی‌های مبتنی بر *DEA* می‌توانند بازده مازاد قابل توجهی را به‌دست آورند [11]. اسکرینجاریک یک مدل پویا *SBM DEA* را برای ارزیابی کارایی نسبی سهام پیشنهاد کرد که در میان آن بازده سه ماهه به‌عنوان یک خروجی، انحراف استاندارد به‌عنوان ورودی و ضرایب چولگی به‌عنوان شرایط پیوند درنظر گرفته می‌شود. یافته‌های تجربی نشان می‌دهد که استراتژی سرمایه‌گذاری مبتنی بر *DEA* پویا می‌تواند از عملکرد بازار بهتر عمل کند [12]. هوآنگ و همکاران از *DEA* برای یافتن سهام باارزش سرمایه‌گذاری استفاده کرد و سپس یک مدل بهینه‌سازی پرتفوی چندهدفه یکپارچه را توسعه داد [13]. گردیجان و اسکرینجاریک انتخاب سهام را با استفاده از امتیازهای کارایی *DEA* آن‌جام دادند، همچنین دو استراتژی سرمایه‌گذاری مبتنی بر *DEA*، از جمله نسبت‌های مساوی و استراتژی‌های وزن‌دهی امتیاز، ساختند. نتایج نشان می‌دهد که استراتژی سرمایه‌گذاری مبتنی بر *DEA* می‌تواند از بازده بازار بهتر عمل کند [14]. شاخص‌های ورودی و خروجی طرح انتخاب سهام مبتنی بر *DEA* در موارد بالا، بیش‌تر از صورت‌های مالی و داده‌های معاملات سهام شرکت‌های فهرست شده مشتق شده‌اند. دوره انتشار اطلاعات مالی شرکت‌های بورسی نسبتا طولانی است که دارای تاخیر زمانی آشکار است. با توجه به شیوع روزافزون معاملات با فرکانس بالا، روش‌های انتخاب سهام و وزن‌دهی پرتفوی بر اساس *DEA* فوق به‌طور قابل توجهی پاسخ‌گو نیستند. در واقع، داده‌هایی که سرمایه‌گذاران می‌توانند مستقیما مشاهده کنند، داده‌های معاملاتی سهام است (به‌عنوان مثال، قیمت افتتاحیه، بالاترین قیمت، پایین‌ترین قیمت، قیمت پایانی و شاخص‌های فنی و غیره).

ساختار باقی‌مانده این مقاله به شرح زیر است. در بخش ۳، ما یک طرح انتخاب سهام را برای تجزیه‌وتحلیل ارزش سرمایه‌گذاری سهام ارایه می‌کنیم. سپس یک مدل بهینه‌سازی پرتفوی را بر اساس طرح انتخاب سهام ارایه شده و حرکات قیمت سهام برگرفته از الگوریتم *SVM* ارایه می‌کنیم. در بخش ۴، ما سهام تشکیل‌دهنده شاخص بورس اوراق‌بهادار تهران را به‌عنوان سهام جایگزین برای ارزیابی طرح انتخاب سهام ارایه شده و استراتژی‌های سرمایه‌گذاری در چارچوب خارج از نمونه درنظر می‌گیریم. در نهایت، نتایج این مقاله را در بخش ۵ نتیجه‌گیری می‌کنیم.

۳- روش‌شناسی پژوهش

۳-۱- گردآوری منابع داده

در فرایند انتخاب پرتفوی، روش موثر انتخاب سهام تا حد زیادی تعیین‌کننده موفقیت استراتژی سرمایه‌گذاری خواهد بود. در عمل، سرمایه‌گذاران معمولا سهام را بر اساس عملکرد تاریخی آن‌ها بررسی می‌کنند. با این حال، برخی از پژوهشگران معتقدند که یک اوراق‌بهادار که در گذشته خوب عمل می‌کند لزوما عملکرد خوب آن را در آینده تضمین نمی‌کند. در واقع، همبستگی بین عملکرد گذشته یک دارایی و عملکرد آتی آن به اثربخشی بازارهای مالی بستگی دارد. واضح است که داده‌های تاریخی معاملات سهام نمی‌توانند به‌طور کامل ارزش سرمایه‌گذاری آتی سهام را منعکس کنند. به این دلایل، برای ارزیابی ارزش سرمایه‌گذاری سهام، داده‌های چندگانه از سهام، از جمله داده‌های معاملاتی تاریخی و شاخص فنی را جمع آوری خواهیم کرد.

این مقاله داده‌های معاملات تاریخی روزانه سهام را جمع‌آوری می‌کند. اطلاعات جمع‌آوری‌شده شامل قیمت افتتاحیه، بالاترین قیمت، پایین‌ترین قیمت، قیمت بسته شدن و حجم معاملات می‌باشد. داده‌های معاملات تاریخی فوق می‌توانند عملکرد تاریخی سهام را منعکس کنند. همچنین می‌توانند راهنمایی استراتژیک متناظری را برای انتخاب سهام سرمایه‌گذاران تا حد معینی ارایه دهند.

۲-۱-۳- شاخص‌های فنی

علاوه بر داده‌های تاریخی معاملات، داده‌های شاخص فنی سهام نیز می‌تواند مرجعی برای انتخاب سهام باشد. با اشاره به کار [15]، [16]، ما شاخص تصادفی[1]، شاخص قدرت نسبی[2] و میانگین متحرک همگرایی و واگرایی[3] را به‌عنوان شاخص‌های ارزیابی فنی انتخاب می‌کنیم. از فرمول محاسبه شاخص‌های فنی [17] می‌توانیم ارزش شاخص‌های فنی را بر اساس داده‌های معاملات تاریخی به‌دست آوریم. فرمول‌های محاسباتی به شرح زیر است.

$$MACD = EMA[Stock\ Prices, 12] - EMA[Stock\ Prices, 26]. \quad (۱)$$

میانگین متحرک همگرا-واگرا

این شاخص ارتباط بین دو *EMA* قیمت‌ها را نشان می‌دهد. این شاخص از تفاوت میانگین متحرک نمایی ۲۶ روزه سهم از میانگین متحرک نمایی ۱۲ روزه قیمت آن برآورد می‌شود. خط سیگنال که همان *EMA* نه روزه است، معیاری است که از مقایسه آن با *MACD* فرصت‌های خرید و فروش شناسایی می‌گردند. بر این اساس حالت فروش زمانی است که *MACD* زیر خط سیگنال قرار می‌گیرد. به‌طور مشابه، یک سیگنال خرید زمانی روی می‌دهد که *MACD* بالای خط سیگنال قرار بگیرد. رابطه آن به شرح **معادله (۱)** است.

$$RSI = 100 - \frac{100}{1 + RS}. \quad (۲)$$

$$RS = \frac{\text{متوسط بازده ها در } n \text{ روزی که قیمت پایانی بیش‌تر از قیمت پایانی روز قبل بوده است}}{\text{متوسط زیان ها در } n \text{ روزی که قیمت پایانی کم‌تر از قیمت پایانی روز قبل بوده است}}. \quad (۳)$$

شاخص قدرت نسبی

شاخص *RSI* مقدار بازده‌های اخیر سهام با زیان‌های اخیر آن را مورد مقایسه قرار می‌دهد و این اطلاعات را به عددی بین ۰ تا ۱۰۰ تبدیل می‌کند. این شاخص را می‌توان از **معادله (۲)** محاسبه نمود. اگر *RSI* از ۳۰ بیش‌تر شود سیگنال فروش و اگر از ۷۰ کم‌تر شود سیگنال خرید صادر می‌شود.

شاخص تصادفی

معمولا برای تجزیه‌وتحلیل روندهای کوتاه‌مدت یا میان‌مدت بازار سهام استفاده می‌شود. از طریق تجزیه‌وتحلیل شاخص *KDJ*، پژوهشگران می‌توانند روند قیمت سهام را پیش‌بینی کنند و از قدرت بازار احتیاط کنند. برای پیش‌بینی حرکت قیمت سهام، شاخص سریع *K* را در شاخص *KDJ* انتخاب می‌کنیم. شاخص *KDJ (%K)* دارای مقداری بین ۰ تا ۱۰۰ است. به‌طور کلی، مقدار بالای ۹۰ بیش از حد خرید و زیر ۱۰ بیش از حد فروخته می‌شود. فرمول محاسبه به شرح **معادله (۴)** است:

$$\%K = \left[\frac{\left(\text{قیمت بسته شدن} - \text{کم‌ترین قیمت}\right)}{\left(\text{بالاترین قیمت} - \text{کم‌ترین قیمت}\right)}\right] \times 100. \quad (۴)$$

۲-۳- طراحی طرح انتخاب سهام

ارزش سرمایه‌گذاری سهام از دو جنبه زیر ارزیابی می‌شود:

---

[1] KDJ
[2] Relative Strength Index (RSI)
[3] Moving Average Convergence Divergence (MACD)

۱. عملکرد بازده تاریخی سهام در حال ارزیابی: بسیاری از پژوهشگران بر این باورند که اثر حرکت خاصی در بازار سهام وجود دارد، یعنی سهامی که در گذشته عملکرد خوبی داشته‌اند، در آینده نیز عملکرد خوبی خواهند داشت؛ بنابراین، این مقاله عمدتا بر تحلیل میانگین و واریانس بازده تاریخی سهام متمرکز است. اگر یک سهم بازده مورد انتظار تاریخی بالاتر و واریانس کم‌تری داشته باشد، نشان می‌دهد که سهام ممکن است دارای ارزش سرمایه‌گذاری مشخصی باشد.

۲. میزان همبستگی بین سهام در حال ارزیابی: هنگام ساخت یک سبد، سرمایه‌گذاران ترجیح می‌دهند سهامی را با همبستگی کم‌تر انتخاب کنند تا بتوانند ریسک‌های سرمایه‌گذاری را به‌طور موثرتری پوشش دهند.

در پژوهش حاضر از مدل تابع فاصله جهت‌دار برای ارزیابی ارزش سرمایه‌گذاری سهام استفاده می‌شود. در مدل *DDF* برای تغییرات ورودی و خروجی جهتی انتخاب می‌شود و واحدها در آن جهت حرکت می‌کنند. ماتریس قیمت‌های تاریخی m سهام درنظر گرفته شده A است، که A یک ماتریس $T\times m$ است و T نشان‌دهنده تعداد روزهای معاملاتی است. بر اساس قیمت‌های تاریخی جمع‌آوری شده (توجه داشته باشید که قیمت‌های بسته‌شدن روزانه سهام در این پژوهش استفاده شده است)، بازده ناخالص روزانه سهام را می‌توان به صورت $\Delta p_{t+1,j} = 1 + \left(p_{t+1,j} - p_{t,j}\right)/p_{t,j}$ بیان کرد. که در آن $p_{t,j}$ و $p_{t+1,j}$ به‌ترتیب نشان‌دهنده قیمت بسته شدن سهام j در بازه‌های زمانی t و $t+1$ است $t = 1,2, \dots , T-1$ و $j = 1,2, \dots , m$.
علاوه‌بر این، فرض می‌کنیم که r_j و σ_j به‌ترتیب میانگین و واریانس نمونه بردار بازده تاریخی $\Delta p_j = \left[\Delta p_{2,j}, \Delta p_{3,j}, \dots , \Delta p_{T,j}\right]'$ را نشان می‌دهند که در آن $j = 1,2, \dots , m$ است. ما از فرض بازده به مقیاس متغیر[1] برای ایجاد **امکان پرتفوی (۵)** که تحت چارچوب میانگین–واریانس کلاسیک تنظیم شده است، استفاده می‌کنیم.

$$P_1 = \left\{(r,\sigma)|r \le \sum_{j=1}^{m} \lambda_j r_j , \sigma \ge \sum_{j=1}^{m} \lambda_j \sigma_j , \sum_{j=1}^{m} \lambda_j = 1, \lambda_j \ge 0, j = 1,2, \dots , m\right\}. \quad (۵)$$

مجموعه **امکان پرتفوی (۵)** از کار مارکوویتز مشتق شده است که می‌تواند به‌عنوان تقریب خطی جزئی آن در مارکوویتز درنظر گرفته شود [18]. ساخت **مجموعه امکان (۵)** توسط سایر پژوهشگران در ارزیابی پرتفوی به رسمیت شناخته شده است [19-24]. بازده به مقیاس بیانگر پیوند بین تغییرات ورودی‌ها و خروجی‌های یک سیستم می‌باشد. در واقع افزایش متناسب بیشینه همه خروجی‌ها به سبب افزایش متناسب معین همه ورودی‌ها است [25]. هنگامی که فقط به رابطه بین معیارهای میانگین و واریانس نگاه می‌کنیم، فرض *VRS* ممکن است به دلیل رابطه غیرخطی بین این دو کاربرد بیش‌تری داشته باشد. تنها زمانی که میانگین و انحراف استاندارد به‌عنوان شاخص‌های ارزیابی استفاده می‌شود، می‌توان فرض بازده به مقیاس ثابت[2] (که معادل خطی بودن است) اتخاذ شود، زیرا میانگین و انحراف استاندارد یک رابطه خطی دارند [25]. علاوه بر این، از نظر تئوری قیمت لذت‌گرا، با توجه به ماهیت غیرخطی عموما در توابع قیمت ضمنی برای ویژگی‌ها، نیاز به تحمیل *VRS* به جای *CRS* است [26]. فرض *CRS* ممکن است در ارزیابی پرتفوی قابل اجرا نباشد، زیرا اجازه می‌دهد بازدهی بیش‌تر از هر پرتفوی تولید کند [27]؛ بنابراین، در مقایسه با مدل *CRS*، معتقدیم که مدل *VRS* ممکن است برای تحلیل ارزش سهام در این پژوهش مناسب‌تر باشد.

اجازه دهید O پانویس سهام مورد ارزیابی را مشخص کند. در ادامه، از تابع فاصله جهت‌دار کلاسیک برای ارزیابی عملکرد سهام استفاده می‌کنیم که می‌تواند ریسک را کاهش و بازده را افزایش دهد. برای یک جهت طرح‌ریزی داده شده $(g_r, -g_\sigma)'$ ، با فرض $(r_0 + \eta g_r, \sigma_0 - \theta g_\sigma) \in P_1$ ، می‌توانیم مدل ارزیابی *DEA* زیر را بسازیم:

$$\theta_1 = \min_{\theta,\eta}\{\frac{1-\theta}{1+\eta} | \; r_0 + \eta g_r, \sigma_0 - \theta g_\sigma) \in P_1\}.$$

به‌طور خاص، می‌توان آن را به‌صورت **مدل (۶)** بیان کرد:

---

[1] Variable Returns to Scale (VRS)
[2] Constant Returns to Scale (CRS)

$$\theta_1 = \min \frac{1-\theta}{1+\eta}.$$
$$s.t.\begin{cases} \sum_{j=1}^{m} \lambda_j r_j \geq r_o + \eta g_r, \\ \sum_{j=1}^{m} \lambda_j \sigma_j \leq \sigma_o - \theta g_\sigma, \\ \sum_{j=1}^{m} \lambda_j = 1, \theta \geq 0, \eta \geq 0\,, \\ \lambda_j \geq 0, j = 1,2, \dots, m. \end{cases} \quad (۶)$$

توجه داشته باشید که، در این‌جا فرض می‌کنیم که جهت $g_r = r_0$ و $g_\sigma = \sigma_0$ را برآورده می‌کند.

از آن‌جایی که تابع هدف **مدل (۶)** یک تابع غیرخطی است، **مدل (۶)** را با تبدیل زیر خطی می‌کنیم. اجازه دهید $1/\tau = 1+\eta, \lambda_j = \tau\lambda_j, \theta = \tau\theta \; and \; \eta = \tau\eta$ شود، سپس **مدل (۶)** را می‌توان به‌صورت **مدل (۷)** بازنویسی کرد.

$$\theta_1 = min\,\tau - \theta.$$
$$s.t.\begin{cases} \sum_{j=1}^{m} \lambda_j r_j \geq \tau r_o + \eta g_r, \\ \sum_{j=1}^{m} \lambda_j \sigma_j \leq \tau\sigma_o - \theta g_\sigma, \\ \tau + \eta = 1, \eta \geq 0, \tau \geq 0, \theta \geq 0, \\ \sum_{j=1}^{m} \lambda_j = \tau, \lambda_j \geq 0, j = 1,2, \dots m. \end{cases} \quad (۷)$$

مدل (۶) و **مدل (۷)** عمدتا کارایی سهام را از منظر میانگین و واریانس بازده ارزیابی می‌کنند؛ با این حال، کار پیشگام مارکوویتز نشان داد که هر چه همبستگی بین سهام کم‌تر باشد، پرتفوی بیش‌تر می‌تواند ریسک را کاهش دهد [2]. بر اساس بازده سهام تحت ارزیابی، ماتریس ضریب همبستگی را می‌توان محاسبه کرد که با $\Theta = \left[\rho_{i,j}\right]_{m\times m}$ مشخص می‌شود. به این معنا که برای دارایی j می‌توانیم دنباله‌ای از ضرایب همبستگی بین بازده این دارایی و سایر دارایی‌های جایگزین را به‌دست آوریم. علاوه بر این، باید دنباله ضریب همبستگی بالا را اندازه‌گیری کنیم تا همبستگی جامع بین دارایی j و سایر دارایی‌ها را توصیف کنیم. به منظور اندازه‌گیری بیش‌تر همبستگی بین سهام انتخاب شده، این پژوهش میانگین و واریانس ضرایب همبستگی در ستون j را به ترتیب ρ_j و σ_j درنظر می‌گیرد که $j = 1,2, \dots, m$ است. سرمایه‌گذاران می‌خواهند ρ_j و σ_j تا حد امکان کوچک باشند، زیرا این می‌تواند اطمینان حاصل کند که ضریب همبستگی بین سهام انتخاب‌شده و سایر سهام دارای میانگین و نوسان کم است. به‌طور مشابه، با استفاده از فرض VRS می‌توانیم **مجموعه امکان پرتفوی (۸)** را بسازیم.

$$P_2 = \left\{ \rho, \sigma) | \rho \geq \sum_{j=1}^{m} \mu_j \rho_j\,, \sigma \geq \sum_{j=1}^{m} \mu_j \sigma_j\,, \sum_{j=1}^{m} \mu_j = 1, \mu_j \geq 0, j = 1,2, \dots, m \right\}. \quad (۸)$$

هر دو میانگین و واریانس در **مجموعه (۸)** دو شاخص ارزیابی هستند که از ماتریس ضریب همبستگی Θ مشتق شده‌اند. در این‌جا، معتقدیم که هر چه دو شاخص فوق کوچک‌تر باشند، بهتر است. بر اساس مجموعه بالا، برای یک جهت طرح‌ریزی داده شده $\left(-g_\rho, -g_\sigma\right)'$، با فرض این‌که $\left(\rho_0 - \alpha g_\rho, \sigma_0 - \beta g_\sigma\right) \in P_2$، می‌توانیم مدل ارزیابی DEA زیر را بسازیم:

$$\theta_2 = \min_{\alpha,\beta} \left\{1 - \frac{\alpha+\beta}{2} | \left(\rho_0 - \alpha g_\rho, \sigma_0 - \beta g_\sigma\right) \in P_2 \right\}.$$

شکل خاص را می‌توان به صورت **معادله (۹)** بیان کرد:

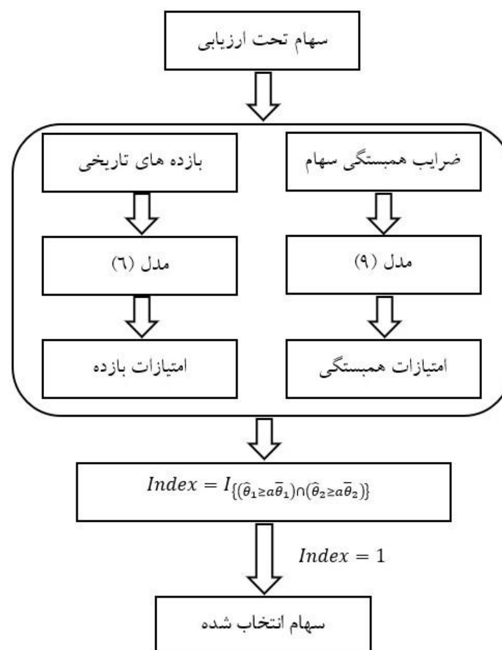
$$\theta_2 = \min 1 - \ \alpha + \beta)/2.$$
$$s.t.\begin{cases}\sum_{j=1}^{m}\mu_j\rho_j \le \rho_o - \alpha g_\rho,\\ \sum_{j=1}^{m}\mu_j\sigma_j \le \sigma_o - \beta g_\sigma,\\ \sum_{j=1}^{m}\mu_j = 1\,, \alpha \ge 0, \beta \ge 0,\\ \mu_j \ge 0, j = 1,2, \dots, m.\end{cases} \quad (۹)$$

در این‌جا، جهت بالا فرض می‌شود که $g_\rho = \rho_0$ و $g_\sigma = \sigma_0$ را برآورده می‌کند زیرا داده‌های مطالعه تجربی همه بزرگ‌تر از ۰ هستند.

بر اساس مدل‌های DEA فوق، می‌توانیم کارایی متناظر سهام را تحت معیارهای ارزیابی مختلف به‌دست آوریم، یعنی θ_1 و θ_2. به‌منظور تسهیل طراحی طرح انتخاب سهام، میانگین کارایی تمام سهام تحت معیارهای ارزیابی مختلف به ترتیب θ_1 و θ_2 تعریف می‌شود. سپس، می‌توانیم شاخص انتخاب سهام زیر را بسازیم.

$$Index = I_{\{(\theta_1 \ge a\theta_1)\cap(\theta_2 \ge a\theta_2)\}}. \quad (۱۰)$$

که در آن $I_{\{.\}}$ یک تابع نشان‌گر است و ضریب a یک سطح از پیش تعیین‌شده بین ۰ و ۱ است. در این‌جا a را می‌توان به‌عنوان اولویت انتخاب سهام تصمیم‌گیرندگان درنظر گرفت. بدیهی است که هر چه ارزش a بزرگ‌تر باشد، الزامات تصمیم‌گیرندگان سخت‌گیرانه‌تر است؛ بنابراین، سهام کم‌تری انتخاب می‌شود. مراحل خاص انتخاب سهام در **شکل ۱** نشان داده شده است.



شکل ۱- جریان انتخاب سهام بر اساس DEA با منابع داده چندگانه.

Figure 1- Stock selection flow based on DEA with multiple data sources.

بر اساس فرایند انتخاب سهام نشان داده شده در **شکل ۱**، سهام مربوطه را می‌توان به‌عنوان دارایی‌های اساسی برای ساخت پرتفوی انتخاب کرد.

۳-۳- پیش‌بینی حرکات آتی قیمت سهام

داده‌های چند منبع فوق نه‌تنها می‌توانند به‌عنوان مبنایی برای انتخاب سهام، بلکه برای پیش‌بینی روند آتی قیمت سهام نیز مورد استفاده قرار گیرند. فراز و نشیب‌های آتی سهام نیز می‌تواند حمایت تصمیم‌گیری مربوطه را برای سرمایه‌گذاران برای ایجاد یک سبد سرمایه‌گذاری موثر فراهم کند.

ایده الگوریتم طبقه‌بندی SVM را به‌طور خلاصه به‌صورت زیر شرح می‌دهیم.

برای مجموعه آموزشی داده شده $A = \{ (\hat{x}_1, \hat{y}_1), (\hat{x}_2, \hat{y}_2), \dots, (\hat{x}_N, \hat{y}_N)\}$، مجموعه ویژگی $\hat{x}_i \in R^{\hat{m}}$ است و آیتم دسته‌بندی $\hat{y}_i = \{0,1\}$ را برآورده می‌کند، جایی که $i = 1,2, \dots, N$ است. در این پژوهش، N نشان‌دهنده طول دوره نمونه (یعنی تعداد روزهای معاملاتی) و $\hat{m}$ نشان‌دهنده مقدار داده‌های آموزشی (یعنی داده‌های معاملات تاریخی و شاخص‌های فنی) در روز معاملاتی i که $i = 1,2, \dots, N$، برای هر سهم در حال ارزیابی است.

تابع نگاشت $\emptyset(\hat{x})$ و همچنین پارامتر جریمه C را انتخاب کنید و جواب‌های بهینه b^*، w^*، ε^* را با توجه به **مساله بهینه‌سازی (۱۱)** به‌دست آورید.

$$\begin{aligned} &\min_{w,b,\varepsilon} \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \varepsilon_i. \\ &s.t. \quad \hat{y}_i\, w^T \varnothing\, \hat{x}_i) + b) + \varepsilon_i \geq 1. \\ &\qquad \varepsilon_i \geq 0, i = 1,2, \dots, N. \end{aligned} \qquad (۱۱)$$

در این‌جا، ε_i عبارت ریسک تجربی است.

تابع کرنل $K(\hat{x}_i, \hat{x}_j) = \varnothing\, \hat{x}_i).\varnothing(\hat{x}_j)$ در مدل‌سازی از تابع کرنل پایه شعاعی استفاده می‌شود.

ابرصفحه متمایز شده $w^*)^T\hat{x} + b^* = 0$ را بر اساس راه‌حل‌های بهینه b^*، w^*، ε^* بسازید تا **تابع تصمیم (۱۲)** را به‌دست آورید:

$$f(\hat{x}) = sgn\,((w^*)^T \emptyset(\hat{x}) + b^*). \qquad (۱۲)$$

توجه داشته باشید که $sgn\ \ z)$ نشان‌دهنده تابع علامت است، زمانی که $z \geq 0$، $sgn\ \ z) = 1$، در غیر این صورت $sgn\ \ z) = 0$.

علاوه بر این، شکل دوگان **مدل (۱۱)** را می‌توان به‌صورت **مدل (۱۳)** بیان کرد.

$$\begin{aligned} &\min_a \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \hat{y}_i\hat{y}_j a_i a_j K(\hat{x}_i, \hat{x}_j) - \sum_{i=1}^{N} a_i. \\ &s.t. \begin{cases} \sum_{i=1}^{N} \hat{y}_i a_i = 0, \\ 0 \leq a_i \leq C, i = 1,2, \dots, N. \end{cases} \end{aligned} \qquad (۱۳)$$

جواب بهینه مدل بالا به صورت $a^* = \ a_1^*, a_2^*, \dots, a_N^*)^T$ ثبت می‌شود، سپس تابع تصمیم نیز می‌تواند به صورت **معادله (۱۴)** بیان شود:

$$f\ \hat{x}) = sgn\left(\sum_{i=1}^{N} \hat{y}_i a_i^* K\ \hat{x}_i, \hat{x}) + b^*\right). \qquad (۱۴)$$

۴-۳- ارایه و ارزیابی مدل بهینه‌سازی پیشنهادی

بر اساس طرح انتخاب سهام پیشنهادی، می‌توانیم گروهی از سهام را از سهام در حال ارزیابی غربال کنیم که می‌توان آن‌ها را به‌عنوان دارایی‌های زیربنایی پرتفوی درنظر گرفت. برای راحتی، فرض می‌کنیم که تعداد سهام انتخاب شده n است و ماتریس بردار میانگین و کوواریانس بازده تاریخی به ترتیب $R_n = \ r_1, r_2, \dots, r_n)'$ و Ω_n نامیده می‌شوند. در ادامه، با ادغام طرح انتخاب سهام فوق، نتایج پیش‌بینی قیمت سهام و نظریه پرتفوی، یک

استراتژی سرمایه‌گذاری ترکیبی برای سرمایه‌گذاران ارایه خواهیم کرد. در این پژوهش، ما نتایج پیش‌بینی قیمت سهام را با تئوری پرتفوی حداقل واریانس سراسری ترکیب می‌کنیم تا برای سرمایه‌گذاران یک استراتژی سرمایه‌گذاری ترکیبی فراهم کنیم. به‌طور خاص، مدل بهینه‌سازی پرتفوی پیشنهادی را می‌توان به صورت **مدل (۱۵)** بیان کرد.

$$
\begin{aligned}
&\min_x x'\Omega_n x - \phi \times x' f_n. \\
&s.t. \sum_{i=1}^{n} z_i = k. \\
&\sum_{i=1}^{n} x_i = 1. \\
&l_i z_i \le x_i \le u_i z_i \quad i = 1, 2, \dots, n). \\
&l_i z_i \le x_i \le u_i z_i \quad i = 1, 2, \dots, n). \\
&x_i \ge 0 \quad i = 1,2, \dots, n).
\end{aligned}
\tag{۱۵}
$$

در این‌جا، $f_n = \left(f_n^1, f_n^2, \dots, f_n^n\right)'$ بردار پیش‌بینی شده قیمت سهام را نشان می‌دهد. علاوه بر این، f_n^i یک متغیر ۰-۱ است، که در آن $f_n^i = 1$ نشان می‌دهد که قیمت سهام در آینده کاهش نخواهد یافت و $f_n^i = 0$ نشان می‌دهد که قیمت سهام در آینده کاهش می‌یابد، $i = 1,2, \dots, n$. ضریب ϕ یک پارامتر از پیش تعیین شده است که می‌تواند به‌عنوان پارامتر معاوضه بین به حداکثر رساندن شاخص پیش‌بینی قیمت سهام و به حداقل رساندن واریانس بازده پرتفوی درنظر گرفته شود. عدد k تعداد سهام موجود در پرتفوی می‌باشد که بیانگر محدودیت کاردینال است. پارامترهای l_i و u_i به ترتیب کف و سقف هر سهم در پرتفوی می‌باشد.

مدل‌های بهینه‌سازی پرتفوی موجود معمولا بر اساس بازده دارایی برای ایجاد هدف بهینه‌سازی هستند و سپس از بازده تاریخی دارایی‌های اساسی برای تخمین پارامترهای ورودی مورد نیاز برای هدف بهینه‌سازی استفاده می‌کنند. با این حال، در واقع، ویژگی‌های آماری بازده‌های تاریخی دارایی ممکن است در آینده نتوانند آن‌ها را به‌دست آورند که باعث می‌شود عملکرد استراتژی سرمایه‌گذاری حاصل رضایت‌بخش نباشد [30-28]. از جمله محدودیت‌های کاربردی در بهینه‌سازی سبد سرمایه‌گذاری می‌توان به حداقل یا حداکثر کردن میزان سرمایه‌گذاری در یک دارایی اشاره نمود که توسط بیل و فارست به مدل اولیه مارکوویتز اضافه شد [31]. با توجه به تمایل سرمایه‌گذاران به در اختیار داشتن و مدیریت سبدی از سهام که تعداد محدودتری از دارایی‌ها را در خود جای داده باشد، محدودیت دیگری تحت عنوان محدودیت کاردینالیتی به مدل اضافه شد. بر خلاف مدل‌های موجود، مدل پیشنهادی دارای دو ویژگی زیر است: الف) دارایی‌های اساسی در **مدل (۱۵)** با استفاده از روش انتخاب سهام در بخش ۳-۲ انتخاب می‌شوند. و ب) **مدل (۱۵)** شاخص‌های پیش‌بینی قیمت سهام را دربر می‌گیرد که به یک مبادله بین عملکرد گذشته و آینده پرتفوی می‌رسد و همچنین وابستگی بیش از حد استراتژی سرمایه‌گذاری منتج به عملکرد تاریخی دارایی‌های اساسی را تا حد زیادی تعدیل می‌کند.

برای نشان دادن اعتبار استراتژی سرمایه‌گذاری پیشنهادی، استراتژی حداقل واریانس سراسری کلاسیک را برای مقایسه ارایه می‌کنیم. برای این منظور، ابتدا فرض می‌کنیم که سرمایه‌گذاران شاخص‌های پیش‌بینی قیمت سهام را درنظر نمی‌گیرند و فقط نگران واریانس بازده پرتفوی هستند و سپس می‌توان مدل بهینه‌سازی پرتفوی حداقل واریانس سراسری متناظر را به صورت **مدل (۱۶)** ساخت.

$$
\begin{aligned}
&\min_x x'\Omega_n x. \\
&s.t. \sum_{i=1}^{n} z_i = k. \\
&\sum_{i=1}^{n} x_i = 1. \\
&l_i z_i \le x_i \le u_i z_i \quad i = 1, 2, \dots, n). \\
&z_i \in [0,1] \quad i = 1, 2, \dots, n). \\
&x_i \ge 0 \quad i = 1,2, \dots, n).
\end{aligned}
\tag{۱۶}
$$

برخی از شاخص‌های ارزیابی کلاسیک (مانند نسبت شارپ و نسبت سورتینو) را برای اندازه‌گیری بازده پرتفوی اعمال می‌کنیم و سپس تفاوت در عملکرد استراتژی‌های سرمایه‌گذاری مختلف را بررسی می‌کنیم.

اجازه دهید دنباله بازده پرتفوی به‌دست آمده از یک استراتژی سرمایه‌گذاری مشخص، یک متغیر تصادفی r_p باشد، سپس نسبت شارپ و نسبت سورتینو مربوط به بازده پرتفوی را می‌توان به ترتیب و به‌صورت **معادله (۱۷)** و **معادله (۱۸)** بیان کرد.

$$z_i \in [0,1] \qquad i = 1, 2, \dots, n). \tag{۱۷}$$

$$\theta^{Sortino} = \frac{\overline{E}(r_p) - r_f}{\sqrt{\overline{E}\left[\left(r_p - \overline{E}(r_p)\right)^2\right]}}. \tag{۱۸}$$

در این‌جا $\overline{E}(.)$ و $\overline{Var}(.)$ به ترتیب میانگین نمونه و واریانس نمونه را نشان می‌دهند. توجه داشته باشید از آن‌جایی که افق سرمایه‌گذاری کوتاه است، نرخ بدون ریسک را ۰ فرض می‌کنیم.

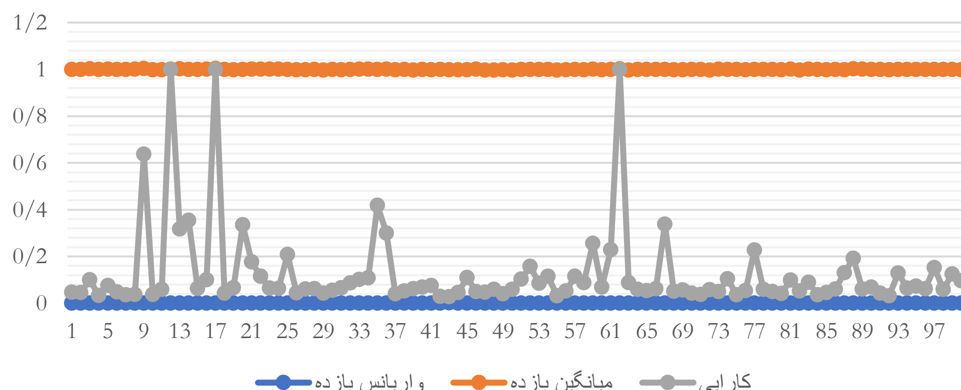
۴- تجزیه‌وتحلیل داده‌ها

در این پژوهش داده‌های جمع‌آوری شده، متعلق به ۱۰۰ شرکت پذیرفته شده در سازمان بورس و اوراق‌بهادار تهران از ۲۴ صنعت مختلف است که در حدفاصل شهریورماه ۱۴۰۰ تا آبان‌ماه ۱۴۰۱ اخذ و تحلیل شده‌اند. برای انتخاب مجموعه سهام، معیارهای زیر اعمال شده است:

۱. شرکت قبل از سال ۱۴۰۰ در بورس اوراق‌بهادار تهران پذیرفته شده باشد و نام آن تا پایان آبان ماه ۱۴۰۱ از فهرست شرکت‌های یاد شده حذف نشده باشد.
۲. سال مالی آن‌ها به پایان اسفندماه هر سال منتهی شود.
۳. در دوره مورد بررسی توقف نماد بیش از ۲ ماه نداشته باشند.
۴. اطلاعات آن‌ها کامل و در دسترس باشد.

سیستم عامل مربوط برای اجرای نرم‌افزارها ویندوز نسخه ۱۱ بوده و محاسبات توسط پردازشگر کامپیوتر قابل حمل ۸ هسته‌ای با فرکانس ۱/۸ گیگاهرتز و حافظه موقت ۸ گیگابایت صورت گرفته است.

در ابتدا کارایی دارایی‌ها از نظر بازده و همبستگی با استفاده از روش تابع فاصله جهت‌دار معرفی شده، ارزیابی می‌شود و دارایی‌های زیربنایی پرتفوی انتخاب می‌گردد.

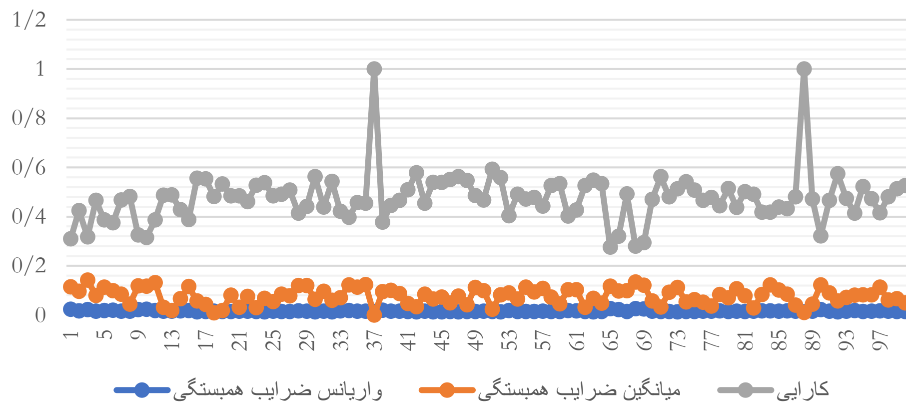


شکل ۲- امتیازات کارایی دارایی‌ها از نظر میانگین-واریانس بازده.

Figure 2- Asset efficiency scores in terms of mean-variance of returns.

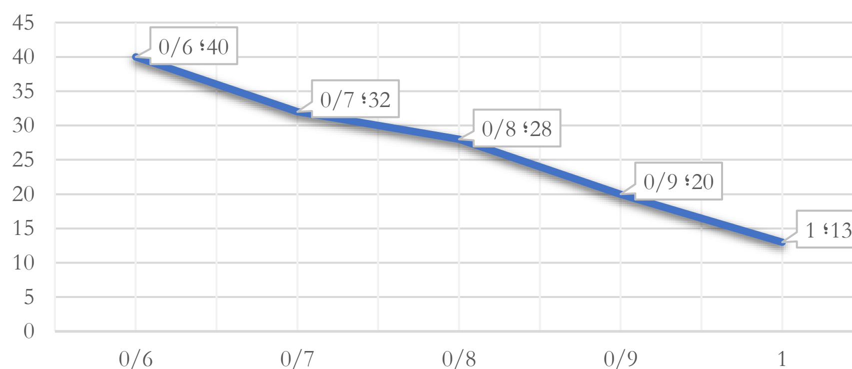
در **شکل ۲** امتیازات کارایی از منظر میانگین-واریانس بازده دارایی‌ها نمایش داده شده است. مطابق شکل سهم‌های ۱۲، ۱۷ و ۶۲ کارا هستند.

به‌منظور اندازه‌گیری همبستگی بین داده‌ها از ضریب همبستگی پیرسون استفاده می‌شود.



شکل ۳- امتیازات کارایی دارایی‌ها از نظر همبستگی بازده.
Figure 3- Asset efficiency scores in terms of return correlation.

در **شکل ۳** امتیازات کارایی از منظر همبستگی بازده دارایی‌ها نمایش داده شده است. مطابق شکل سهم‌های ۳۷ و ۸۸ کارا هستند.



شکل ۴- تعداد سهام انتخابی در سطح اولویت انتخاب سهام a.
Figure 4- Number of selected stocks at stock selection priority level a.

هر چه ارزش a بزرگ‌تر باشد، الزامات تصمیم‌گیرندگان سخت‌گیرانه‌تر است؛ بنابراین، سهام کم‌تری انتخاب می‌شود. **شکل ۴** تعداد سهام انتخابی در سطوح مختلف اولویت انتخاب سهام را نمایش می‌دهد. با درنظر گرفتن سطح اولویت انتخاب سهام ۰/۶ دارایی‌های زیربنایی پرتفوی به شرح **جدول ۱** انتخاب می‌شوند.

جدول ۱- سهام‌های منتخب در سطح اولویت انتخاب سهام $a = 0.6$.
Table 1- Selected stocks at stock selection priority level $a = 0.6$.

j	Name	j	Name	j	Name	j	Name
3	ختوقا	22	شپنا	52	دالبر	74	فولاد
5	خگستر	25	شنفت	53	دتماد	77	کنور
9	خزامیا	32	شستا	54	دفارا	81	غشهد
12	مارون	33	وغدیر	57	دپارس	83	غگل
13	بوعلی	34	وبانک	58	درازک	87	پارسیان
14	جم	35	وامید	59	دسینا	88	دانا
16	حکشتی	36	وصندوق	61	وخاور	93	قشهد
17	حفارس	41	فاما	62	وکار	97	سیتا
20	همراه	45	واعتبار	63	وپارس	99	سرود
21	شبندر	51	دابور	67	وپاسار	100	سدشت

در گام بعد از ماشین بردار پشتیبان[1] ترکیب شده با داده‌های چندمنبعی برای پیش‌بینی حرکات قیمت سهام استفاده می‌کنیم. ابتدا از روش *z_Score* برای نرمال‌سازی مجموعه داده‌های ورودی استفاده می‌کنیم. سپس از *SVM* برای تنظیم، آموزش و پیش‌بینی داده‌های ورودی استفاده می‌شود. ما از روش *Cross-Validation* با *K* برابر ۵ برای یادگیری مدل استفاده می‌کنیم. سپس می‌توانیم هایپرپارامترها را با توجه به نتایج مجموعه تست تنظیم کنیم. به‌طور کلی، زمانی‌که تابع کرنل پایه شعاعی[2] انتخاب می‌شود، عملکرد مدل نسبتا پایدار است؛ بنابراین، در این پژوهش ابتدا تابع کرنل *RBF* انتخاب شده، سپس با استفاده از جست‌وجوی تصادفی بازه کوچک‌تری از مقادیر هایپرپارامترها را فراهم می‌کنیم و فضای جست‌وجو را کوچک‌تر می‌سازیم. در آخر از روش جست‌وجوی شبکه‌ای برای تنظیم هایپرپارامترهای C و γ استفاده می‌شود. محدوده مقدار هایپرپارامتر جریمه C را روی [1,10] قرار می‌دهیم و طول گام را ۱ تعیین می‌کنیم. محدوده مقدار γ را روی [0.001,0.1] قرار می‌دهیم و طول گام را ۰٫۰۰۱ تعیین می‌کنیم. با تنظیم هایپرپارامترها می‌توان هایپرپارامترهای دارای بالاترین دقت را استخراج کرد. تنظیمات هایپرپارامترهای الگوریتم *SVM* در **جدول ۲** نشان داده شده است.

جدول ۲- هایپرپارامترهای تنظیم‌شده الگوریتم به ازای تابع کرنل RBF.
Table 2- Tuned hyperparameters of the algorithm for the RBF kernel function.

Gamma	C	j	Gamma	C	j
0.01	1	52	0.02	3	3
0.014	7	53	0.006	1	5
0.008	9	54	0.007	1	9
0.002	8	57	0.039	9	12
0.005	2	58	0.009	5	13
0.019	3	59	0.001	1	14
0.015	1	61	0.005	1	16
0.011	9	62	0.011	1	17
0.008	3	63	0.003	9	20
0.005	1	67	0.031	5	21
0.002	5	74	0.003	3	22
0.002	9	77	0.01	2	25
0.004	3	81	0.016	4	32
0.01	1	83	0.02	1	33
0.008	9	87	0.008	9	34
0.006	7	88	0.001	7	35
0.001	7	93	0.022	7	36
0.019	8	97	0.006	3	41
0.049	8	99	0.001	9	45
0.147	4	100	0.002	8	51

در ادامه، اثرات پیش‌بینی مدل *SVM* را بر اساس شاخص‌های ارزیابی دقت، صحت، فراخوان، امتیاز *F1* ارزیابی می‌کنیم. در این‌جا، دقت به نسبت نمونه‌های پیش‌بینی شده درست به همه نمونه‌ها اشاره دارد. صحت نشان‌دهنده نسبت نمونه‌های مثبت پیش‌بینی شده صحیح به نمونه‌های مثبت پیش‌بینی شده است. فراخوان نشان‌دهنده نسبت نمونه‌های مثبت پیش‌بینی شده صحیح به نمونه‌های مثبت واقعی است. امتیاز *F1* میانگین هارمونیک صحت و فراخوان است. فرمول‌های محاسباتی به شرح زیر است.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (۱۹)$$

$$Precision = \frac{TP}{TP + FP} \quad (۲۰)$$

$$Recall = \frac{TP}{TP + FN} \quad (۲۱)$$

$$F1_score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (۲۲)$$

[1] Support Vector Machine (SVM)
[2] Radial Basis Function (RBF)

TP نشان‌دهنده مثبت واقعی، *FP* نشان‌دهنده مثبت کاذب، *TN* نشان‌دهنده منفی واقعی و *FN* نشان‌دهنده منفی کاذب است. جزئیات *TP*، *FP*، *TN* و *FN* در **جدول ۳** نشان داده شده است.

جدول ۳- معانی TP، FP، TN و FN.
Table 3- Meanings of TP, FP, TN and FN.

مقدار واقعی \ مقدار پیش‌بینی شده	1	0
1	TP	FN
0	FP	TN

چهار شاخص ارزیابی (دقت، صحت، فراخوان و امتیاز *F1*) بر اساس روش *Cross-Validation* برای ارزیابی عملکرد مدل استفاده می‌شود. عملکرد پیش‌بینی قیمت سهام در **جدول ۴** نشان داده شده است. می‌توان دریافت که میانگین دقت، صحت، فراخوان و امتیاز *F1* مجموعه تست همه سهام به ترتیب ٪۶۷، ٪۶۹، ٪۵۹ و ٪۶۱ است. علاوه بر این، ۳۶ سهام با دقت بیش از ٪۶۰ و ۴۰ سهام با دقت بیش از ٪۵۰ وجود دارد. ۳۸ سهام با صحت بیش از ٪۶۰ و ۴۰ سهام با صحت بیش از ٪۵۰ وجود دارد. ۱۹ سهام با فراخوان بالای ٪۶۰ و ۲۹ سهام با فراخوان بالای ٪۵۰ وجود دارد. ۲۱ سهام با امتیاز *F1* بالای ٪۶۰ و ۳۳ سهام با امتیاز *F1* بالای ٪۵۰ وجود دارد.

جدول ۴- ارزیابی عملکرد پیش‌بینی قیمت سهام.
Table 4- Evaluation of stock price forecasting performance.

Cross-validation شاخص‌ها	حداکثر	حداقل	متوسط	0.6<	0.5<
Accuracy	0.83	0.55	0.67	36	40
Precision	0.87	0.54	0.69	38	40
Recall	0.97	0.37	0.59	19	29
F1-Score	0.89	0.37	0.61	21	33

جدول ۴ نشان می‌دهد که نتایج پیش‌بینی اکثر سهام در محدوده قابل قبولی است؛ بنابراین، می‌توان از این نتایج به‌عنوان مرجعی برای سرمایه‌گذاران به‌منظور انتخاب دارایی‌های اساسی یک پرتفوی استفاده کرد. نتایج پیش‌بینی حرکات قیمت سهام مطابق **جدول ۵** است.

جدول ۵- نتایج پیش‌بینی قیمت سهام.
Table 5- Stock price prediction results.

j	Predict	j	Predict	j	Predict	j	Predict
3	0	22	0	52	0	74	0
5	1	25	1	53	1	77	1
9	0	32	0	54	0	81	1
12	1	33	1	57	0	83	0
13	0	34	1	58	0	87	0
14	0	35	0	59	1	88	1
16	0	36	1	61	0	93	0
17	0	41	0	62	1	97	1
20	1	45	0	63	1	99	1
21	0	51	1	67	1	100	0

در آخر، نتایج پیش‌بینی قیمت سهام را با نظریه پرتفوی حداقل واریانس سراسری ترکیب می‌کنیم تا برای سرمایه‌گذاران یک استراتژی سرمایه‌گذاری ترکیبی ۱ فراهم کنیم. برای نشان دادن اعتبار استراتژی سرمایه‌گذاری پیشنهادی، استراتژی سرمایه‌گذاری حداقل واریانس سراسری کلاسیک ۲ را برای مقایسه ارایه می‌کنیم. افزایش تعداد دارایی‌ها در فرایند تشکیل سبد سرمایه‌گذاری، درنظر گرفتن محدودیت سرمایه‌گذاری در هر یک از دارایی‌ها و نیز لحاظ نمودن محدودیت کاردینالی موضوع بهینه‌سازی سبد سرمایه‌گذاری را به یک مساله *NP-Hard* تبدیل می‌نماید. در نتیجه برای حل مدل از الگوریتم ژنتیک استفاده شده است. نتایج حاصل از اجرای استراتژی‌های سرمایه‌گذاری ۱ و ۲ در **جدول ۶** آورده شده است.

جدول ۶- پرتفوی بهینه استراتژی‌های سرمایه‌گذاری.
Table 6- Optimal portfolio of investment strategies.

استراتژی سرمایه‌گذاری ۱ ($a = 0.9$)

12	36	59	62	77
0.3	0.172318	0.148537	0.236434	0.142711
$a = 0.6$				
12	20	36	51	81
0.3	0.255804	0.244196	0.1	0.1
استراتژی سرمایه‌گذاری ۲ ($a = 0.9$)				
12	13	20	22	36
0.3	0.1	0.281168	0.1	0.218832
$a = 0.6$				
9	12	13	14	100
0.1	0.3	0.152031	0.3	0.147969

شاخص‌های ارزیابی نسبت شارپ و نسبت سورتینو را برای اندازه‌گیری بازده پرتفوی اعمال می‌کنیم و سپس تفاوت در عملکرد استراتژی‌های سرمایه‌گذاری مختلف را بررسی می‌کنیم. زمانی که کارایی با اولویت انتخاب سهام به ترتیب در ۰٫۹ و ۰٫۶ ارزیابی شود، نسبت شارپ و نسبت سورتینو بازده پرتفوی در استراتژی‌های ۱ و ۲ در **جدول ۷** نشان داده شده است.

جدول ۷- عملکرد استراتژی‌های سرمایه‌گذاری.

Table 7- Performance of investment strategies.

استراتژی سرمایه‌گذاری	$a = 0.9$	$a = 0.6$
پنل A: نسبت شارپ بازده پرتفوی		
1	0.846691	0.373045
2	0.666201	0.243218
پنل B: نسبت سورتینو بازده پرتفوی		
1	0.946629	0.417076
2	0.744835	0.271926

برای یک ضریب معین a، استراتژی سرمایه‌گذاری پیشنهادی ۱ نسبت شارپ و سورتینو بالاتری نسبت به استراتژی سرمایه‌گذاری ۲ دارد. برای یک استراتژی سرمایه‌گذاری معین، نسبت شارپ و سورتینو قبل از کاهش با افزایش ضریب انتخاب سهام a تمایل به افزایش دارد. این همچنین نشان می‌دهد که در سرمایه‌گذاری واقعی، دارایی‌های زیربنایی پرتفوی به اندازه بهتر نیست و سرمایه‌گذاران باید تعداد کمی سهام با ارزش سرمایه‌گذاری را از تعداد زیادی دارایی انتخاب کنند. بدیهی است که نتایج فوق همچنین تایید می‌کند که طرح انتخاب سهام پیشنهادی و مدل بهینه‌سازی پرتفوی بر اساس منابع داده‌های چندگانه می‌تواند عملکرد پرتفوی را بهبود بخشد.

۵- بحث و نتیجه‌گیری

هدف این پژوهش، بهینه‌سازی پرتفوی با استفاده از ادغام *DEA* با منابع داده چندگانه با رویکرد یادگیری ماشین بوده است. به‌منظور دستیابی به این هدف، ابتدا مدل تحلیل پوششی داده‌ها را جهت ارزیابی کارایی سهام‌های تحت بررسی از نظر بازده تاریخی و همبستگی دارایی و انتخاب دارایی‌های زیربنایی پرتفوی طراحی نموده‌ایم. پس از مشخص شدن سهام‌های دارای ارزش سرمایه‌گذاری، از ماشین بردار پشتیبان ترکیب شده با داده‌های چندمنبعی برای پیش‌بینی حرکات قیمت سهام استفاده شده است. همچنین روش جست‌وجوی تصادفی و شبکه‌ای را برای تنظیم هایپرپارامترهای الگوریتم ارایه شده مدنظر قرار داده‌ایم. سپس حرکات قیمت سهام و طرح انتخاب سهام پیشنهادی را برای ساخت مدل بهینه‌سازی پرتفوی ترکیب کرده و استراتژی سرمایه‌گذاری مربوطه را برای سرمایه‌گذاران ارایه نمودیم. در مطالعه تجربی، این پژوهش دارایی‌های زیربنایی شاخص بورس اوراق‌بهادار تهران را به‌عنوان نمونه‌های آزمایشی انتخاب نموده و سپس یک آزمون را در مورد طرح انتخاب سهام پیشنهادی و استراتژی‌های سرمایه‌گذاری انجام داده است. نتایج تجربی نشان‌دهنده این بود که طرح انتخاب سهام پیشنهادی می‌تواند عملکرد استراتژی‌های سرمایه‌گذاری مختلف را بهبود بخشد. علاوه بر آن، نسبت شارپ و سورتینو بازده استراتژی پیشنهادی نیز از استراتژی سرمایه‌گذاری حداقل واریانس سراسری به مراتب بهتر است. این نتایج نشان داد که سرمایه‌گذاران باید یک استراتژی سرمایه‌گذاری کم‌تنوع را به جای استراتژی سنتی کاملا متنوع اتخاذ کنند.

در پایان و به‌عنوان پیشنهادها جهت پژوهش‌های آتی، گفتنی است که به‌کارگیری روش‌های دیگر همچون استفاده از اوزان فازی و به‌دست آوردن داده‌های فازی در مدل *FUZZY DEA*، بررسی روابط بین شاخص‌ها و مقایسه نتایج حاصل از به‌کارگیری این روش‌ها، استفاده از ماشین بردار پشتیبان با چند خروجی جهت پیش‌بینی و مقایسه نتایج آن با مدل ارایه شده و چگونگی یافتن اندازه بهینه موجودی سهام که طی آن می‌توان از سایر نظریه‌های پرتفوی و روش‌های حل از قبیل فراابتکاری‌ها استفاده کرد؛ می‌تواند عملکرد استراتژی سرمایه‌گذاری را بهتر نموده و در راستای نزدیکی هر چه بیش‌تر به شرایط واقعی بازار گام برداشت

منابع

[1] Markowitz, H. (2009). Portfolio selection. In *Harry markowitz: Selected works* (pp. 15–30). https://doi.org/10.2307/2975974

[2] Markowitz, H. (1952). Portfolio selection. *The journal of finance*, *7*(1), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

[3] Keynes, J. M., Johnson, E., Moggridge, D. (1983). Keynes as an investor. In *The collected writings of john maynard keynes* (pp. 106–123). https://b2n.ir/rq4766

[4] Uppal, R., & Wang, T. (2003). Model misspecification and underdiversification. *The journal of finance*, *58*(6), 2465–2486. https://doi.org/10.1046/j.1540-6261.2003.00612.x

[5] Boyle, P., Garlappi, L., Uppal, R., & Wang, T. (2011). Keynes meets markowitz: The trade-off between familiarity and diversification. *Management science*, *58*(2), 253–272. https://doi.org/10.1287/mnsc.1110.1349

[6] Liu, H. (2014). Solvency constraint, underdiversification, and idiosyncratic risks. *Journal of financial and quantitative analysis*, *49*(2), 409–430. https://doi.org/10.1017/S0022109014000271

[7] Guidolin, M., & Liu, H. (2016). Ambiguity aversion and underdiversification. *Journal of financial and quantitative analysis*, *51*(4), 1297–1323. https://doi.org/10.1017/S0022109016000466

[8] Florentsen, B., Nielsson, U., Raahauge, P., & Rangvid, J. (2019). The aggregate cost of equity underdiversification. *Financial review*, *54*(4), 833–856. https://doi.org/10.1111/fire.12212

[9] Amiri, M., Darestani Farahani, A., & Mahboob-Ghodsi, M. (2017). *Multi-criteria decision making. Second edition.* , Kian University Press.

[10] Edirisinghe, N. C. P., & Zhang, X. (2007). Generalized DEA model of fundamental analysis and its application to portfolio optimization. *Journal of banking & finance*, *31*(11), 3311–3335. https://doi.org/10.1016/j.jbankfin.2007.04.008

[11] Chen, H. (2008). Stock selection using data envelopment analysis. *Industrial management & data systems*, *108*(9), 1255–1268. https://doi.org/10.1108/02635570810914928

[12] Škrinjarić, T. (2014). Investment strategy on the Zagreb stock exchange based on dynamic DEA. *Croatian economic survey*, *16*(1), 129–160. https://doi.org/10.15179/ces.16.1.5

[13] Huang, C. Y., Chiou, C. C., Wu, T. H., & Yang, S. C. (2015). An integrated DEA-MODM methodology for portfolio optimization. *Operational research*, *15*(1), 115–136. https://doi.org/10.1007/s12351-014-0164-7

[14] Gardijan, M., & Škrinjarić, T. (2015). Equity portfolio optimization: A DEA based methodology applied to the Zagreb Stock Exchange. *Croatian operational research review*, *6*(2), 405–417. https://doi.org/10.17535/crorr.2015.0031

[15] Weng, B., Ahmed, M. A., & Megahed, F. M. (2017). Stock market one-day ahead movement prediction using disparate data sources. *Expert systems with applications*, *79*, 153–163. https://doi.org/10.1016/j.eswa.2017.02.041

[16] Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The north american journal of economics and finance*, *47*, 552–567. https://doi.org/10.1016/j.najef.2018.06.013

[17] Mohammadi, S. (2004). Technical analysis on the Tehran stock exchange. *Financial research*, *6*(1), 97–129. https://dor.isc.ac/dor/20.1001.1.10248153.1383.6.1.5.5

[18] Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the royal statistical society: series c (Applied statistics)*, *29*(2), 119–127. https://doi.org/10.2307/2986296

[19] Branda, M. (2015). Diversification-consistent data envelopment analysis based on directional-distance measures. *Omega*, *52*, 65–76. https://doi.org/10.1016/j.omega.2014.11.004

[20] Choi, H.-S., & Min, D. (2017). Efficiency of well-diversified portfolios: Evidence from data envelopment analysis. *Omega*, *73*, 104–113. https://doi.org/10.1016/j.omega.2016.12.008

[21] Lin, R., & Li, Z. (2020). Directional distance based diversification super-efficiency DEA models for mutual funds. *Omega*, *97*, 102096. https://doi.org/10.1016/j.omega.2019.08.003

[22] Xiao, H., Ren, T., & Ren, T. (2020). Estimation of fuzzy portfolio efficiency via an improved DEA approach. *INFOR: Information systems and operational research*, *58*(3), 478–510. https://doi.org/10.1080/03155986.2020.1734904

[23] Liu, W., Zhou, Z., Liu, D., & Xiao, H. (2015). Estimation of portfolio efficiency via DEA. *Omega*, *52*, 107–118. https://doi.org/10.1016/j.omega.2014.11.006

[24] Zhou, Z., Jin, Q., Xiao, H., Wu, Q., & Liu, W. (2018). Estimation of cardinality constrained portfolio efficiency via segmented DEA. *Omega*, *76*, 28–37. https://doi.org/10.1016/j.omega.2017.03.006

[25] Hosseinzadeh Lotfi, F., Jahanshahlo, G., & Nikoumaram., H. (2011). *Data envelopment analysis and its applications*. *Islamic Azad University, science and research branch, Nafis works*. (**In Persian**). https://b2n.ir/bt2998

[26] Kerstens, K., Mounir, A., & de Woestyne, I. Van. (2011). Non-parametric frontier estimates of mutual fund performance using C- and L-moments: Some specification tests. *Journal of banking & finance*, *35*(5), 1190–1201. https://doi.org/10.1016/j.jbankfin.2010.09.030

[27] Lamb, J. D., & Tee, K.-H. (2012). Data envelopment analysis models of investment funds. *European journal of operational research*, *216*(3), 687–696. https://doi.org/10.1016/j.ejor.2011.08.019

[28] Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'Optimized' optimal? *Financial analysts journal*, *45*(1), 31–42. https://doi.org/10.2469/faj.v45.n1.31

[29] DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The review of financial studies*, *22*(5), 1915–1953. https://doi.org/10.1093/rfs/hhm075

[30] Tu, J., & Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of financial economics*, *99*(1), 204–215. https://doi.org/10.1016/j.jfineco.2010.08.013

[31] Beale, E. M. L., & Forrest, J. J. H. (1976). Global optimization using special ordered sets. *Mathematical programming*, *10*(1), 52–69. https://doi.org/10.1007/BF01580653